

PTO 04-3602

CY=JA DATE=19920527 KIND=A
PN=04-153878

PRE-EDITING SUPPORT PROCESSOR FOR A MACHINE TRANSLATION DEVICE
[Kikai Honyaku Sochi niokeru Zenhenshu Shien Shori Sochi]

Jun Ibuki, et al.

UNITED STATES PATENT AND TRADEMARK OFFICE
Washington, D.C. June 2004

Translated by: FLS, Inc.

PUBLICATION COUNTRY	(19):	JP
DOCUMENT NUMBER	(11):	4-153878
DOCUMENT KIND	(12):	A
PUBLICATION DATE	(43):	19920527
APPLICATION NUMBER	(21):	2-280336
APPLICATION DATE	(22):	19901018
INTERNATIONAL CLASSIFICATION	(51):	G06F 15/38
INVENTOR	(72):	IBUKI, JUN; NISHINO, FUMITO; NAKAMURA, NAOTO; SHIOUCHI, MASATOSHI; FUJI, HIDE
APPLICANT	(71):	FUJITSU CO., LTD.
TITLE	(54):	PRE-EDITING SUPPORT PROCESSOR FOR A MACHINE TRANSLATION DEVICE
FOREIGN TITLE	(54A):	KIKAI HONYAKU SOCHI NIOKERU ZENHENSU SHIEN SHORI SOCHI

1. Title of the Invention

PRE-EDITING SUPPORT PROCESSOR FOR A MACHINE TRANSLATION DEVICE

2. Claims

(1) A pre-editing support processor for a machine translation device for an input document (1) where a sentence structure analyzer (2) conducts structural analysis of input documents (1) and for individual sentences, the sentence analyzer (3) analyzes sentences by determining the parts of speech for each word;

and is comprised of a structure processor (4) that rewrites sentences with a structure that is clearer according to previously established sentence patterns in the text based on the results of the sentence analysis conducted by the aforementioned sentence analyzer (3);

and a display (5) that displays the sentences rewritten based on the results performed by the aforementioned structure processor (4);

and these sentences rewritten with clearer structure are then subject to machine translation processing.

(2) A pre-editing support processor for a machine translation device as claimed in Claim (1) comprised of the structure processor (4) that converts the present and past tense of the verb based on sentence patterns containing such present and past verb tenses.

(3) A pre-editing support processor for a machine translation device as claimed in Claim (1) comprised of the structure processor (4) that is equipped with an assignment unit (44) to output evaluation values for specific word strings in the text and with a control unit

* Numbers in the margin indicate pagination in the foreign text.

(46) to indicate whether or not to divide the text based on the evaluation values.

(4) A pre-editing support processor for a machine translation device as claimed in Claim (1) comprised of the structure processor (4) that separates the text based on the sentence structure found in the text, and that constructs phrases for each or any of the separated sentences.

3. Detailed Explanation of the Invention

/508

[Summary]

This invention relates to a pre-editing support processor for a machine translation device that supports the user with machine translation devices. This has the objective of rewriting text in a form that is not ambiguous to the user when the original sentence is ambiguous, or when there is the possibility that multiple translations could be obtained.

This is comprised of a structure processor that rewrites sentences in a form that is clearer according to previously established sentence patterns in the text based on the results of the sentence analysis conducted by the sentence analyzer.

[Commercial Field of Use]

This relates to a pre-editing support processor for a machine translation device that supports the user with machine translation devices.

With existing machine translation devices, it is difficult to analyze sentence structure without making mistakes due to insufficient context or impartiality of the description. Therefore, it is necessary to conduct processing by determining the true meaning in order to

obtain a correct translation.

[Existing Technology]

Currently it takes a tremendous amount of time for the pre-editing process, which hinders improving the efficiency of the machine translation, for the following reasons.

- 1) It takes time to rewrite so the editor efficiency is poor.
- 2) Since the contents of the translation system processing isn't clearly stated, it is difficult to determine how to edit the original text for translation.

Particularly with long sentences, the translation success rate is extremely poor for sentences not subject to pre-editing. Even if translated successfully, many times the translation output by the machine translation system is extremely hard to read. In this case, the human operator collects the useable parts from the system output and manually builds the translation, which lowers the operating efficiency.

At the present time, to obtain a correct translation from a machine translation device, the user must rewrite the text in a form that can be properly interpreted by the system. It takes a tremendous amount of time for someone to become familiar with the system processing, which dramatically increases the overall cost of machine translation.

An automatic parsing system that can automatically conduct such processing has already been developed but there are many problems with the processing accuracy and it is not practical.

This does not address the concerns about sentences that are easy to read, which increases the amount of effort required to check the

translation and is a factor in the increased cost of translation.

[Problems this Invention is to Solve]

When beginners are using the system, it is not easy to understand how to edit the document targeted for translation to obtain the correct translation effect. Thus there are problems with errors in the translation due to text being input into the system without first being edited. Pre-editing by a person familiar with the logic of the system processing generally requires a complete rewrite, which is extremely inefficient. Then, the results of the machine translation is not always easy to read and it is either hard to understand the translation or it is hard to obtain a final edited translation.

This invention the objective of rewriting text in a form that is not ambiguous to the user when the original sentence is ambiguous, or when there is the possibility that multiple translations could be obtained.

[Means of Solving these Problems]

Figure 1 shows the fundamental structure of this invention. In the figure, 1 refers to the document input for translation. /509
2 is the sentence structure analyzer, that performs analysis of the sentence structure by separating the document into the title, paragraph 1, paragraph 2, 3 is the sentence analyzer that analyzes the input sentence. 4 is the structure processor that rewrites the sentences subject to analysis into a form with a structure that is clear. 5 is the display for the rewritten results. 6 is the separate analysis results search part for other analyses obtained from the sentence analyzer 3.

With this invention, if the original text for machine translation is ambiguous, the original text is rewritten for analysis so the beginner can obtain a correct translation using a system that can edit the original text into several potential rewrites for simplified selection. Also, with succinct sentence structure, sentences can be rewritten so translations can be easy to read.

[Operation]

The sentence structure analyzer 2 analyzes information relating to the structure of the entire sentence input and transmits this and the input sentence to the sentence analyzer 3. The sentence analyzer 3 analyzes the sentence input and sends part of the results of such analysis to the structure processor 4. The structure processor 4 takes the results of such analysis and converts the sentences into a form where the structure is clear to be shown on the display 5. The user then looks at the display and accepts or rejects the results of the write according to whether or not they correspond to their own interpretation.

If rejected, the system transmits other analytical results from the sentence analyzer 3 and processing is again conducted by the structure processor 4. If accepted, the analytical results are output without being changed. The user is authorized to determine how ambiguous sections are to be interpreted for clarity.

Naturally, in these cases, processing can be conducted with the objective of eliminating translation ambiguities. If the purpose is to simply improve the ease of reading the translation, the separate analysis results search on sentence analyzer 6 can be employed for such.

[Embodiment Examples]

Figure 2 shows an embodiment example of the structure processor. Figure 3 shows an example of such processing. Symbol 4 in Fig. 2 is the structure processor, 41 is the word string analysis part and 42 refers to the structure processor.

The word string analysis part 41 analyzes the words of the input sentence and determines the part of speech for each word. The structure processor 42 extracts pre-established sentence patterns and converts it to a clearer sentence with these sentence patterns.

The processing example shown in Fig. 3 is a sentence with the sentence pattern of:

Transitive verb form + particle "te"

That is converted to a form of:

Conclusive verb form + "." + "and". With processing example 1, "erase" is an irregular conjugation and the sentence "Erase the file to finish processing" is converted to "Erase the file. And processing is finished".

With processing example 2, "delete" is an irregular conjugation and the sentence "Delete the file to finish processing" is converted to "Delete the file. And processing is finished".

If there multiple verbs such as "do A, do B and do C", this can be converted to "Do A. And do B. And do C".

Figure 4 shows another embodiment example of the structure processor. Figure 5 shows the processing example. In Fig. 4, symbol 4 is the structure processor, 43 is the form analyzer, 44 is the scanning specific words and assigning evaluation value part (hereafter abbreviated as evaluation value part), 45 is the structure processor

and 46 is the structure processor controller.

In Fig. 4, the evaluation value part 44 finds sentences containing for example, more than 3 predicates based on the word string obtained. If the limit is 3, the structure processor 4 /510 converts it to a form where there are no more than 2 predicates in a sentence. The evaluation value part 44 relays the evaluation value "number of predicates" to the structure processor controller 46 and the structure processor controller 46 checks whether or not the "more than 3 predicates" condition is met and sends the "on/off control signal" to the structure processor 45 to indicate if the sentence should be converted.

The processing example shown in Fig. 5 illustrates "conduct the sentence separating process" if there are "more than 3 predicates".

For the example (1) in the figure, the sentence "Erase the file to finish processing" contains 2 predicates so the sentence separating process is not conducted. For the example (2), the sentence "Extract the data and delete the file to finish processing" contains 3 predicates so the sentence separating process is conducted. In this example, "extract" and "delete" are irregular conjugations and so the sentence is converted to "Extract the data. Delete the file. And finish processing".

Figure 6 shows another embodiment example of the structure processor. Figure 7 is an example of the structure using the sentence structure information while Figure 8 and 9 show examples of the results of processing.

In Fig. 6, symbol 4 is the structure processor, 43 is the form analyzer, 47 is the sentence structure analyzer and 48 is the

structure processing unit.

As described referencing Fig. 7, in Fig. 6 the sentence structure analyzer 47 clarifies the layers of the sentence so the structure processing unit 48 can convert the original text.

Figure 7 (A) shows the format where the sentence structure is clarified for the original sentence "The user deletes the file to finish processing". In the figure, s is the sentence, vp is the predicate and pp is the phrase. For (A) in Fig. 7, (i) "deletes the file" and (ii) "finish processing" correspond to the same predicate "the user". As found in (B) and (C) of Fig. 7, it is separated into two sentences and as shown in Fig. 7(C), instead of "the user" as found in Fig. 7(B), "he" is used.

Therefore, the processing example in Fig. 8 takes the sentence: "The user deletes the file to finish processing" and converts it into:

"The user deletes the file."

"And he finishes processing."

As found in Fig. 9, it is converted in the same manner as in Fig. 8, but the parts of speech for the words are shown.

[Effect of this Invention]

As shown in the description above, this invention enables the proper translation of ambiguous sentences to be selected if there could be multiple translations or if the user doesn't know the target word. Also, this invention enables users who do not possess knowledge of the sentences that can be processed by the machine translation system to select the proper translation if the target word isn't known.

This invention is not limited to any of the examples given in Fig. 2, Fig. 4 or Fig. 6 but can include a combination of any of these.

4. Brief Description of the Diagrams

Figure 1 shows the fundamental structure of this invention. Figure 2 shows one embodiment example for the structure processor. Figure 3 is a processing example for that shown in Fig. 2. Figure 4 shows another embodiment example for the structure processor. Figure 5 is a processing example for that shown in Fig. 4. Figure 6 shows another embodiment example for the structure processor. Figure 7 is an example of the structure using sentence structure information. Figure 8 and Figure 9 show examples for various processing.

In the figures: 1 is the document input, 2 is the sentence structure analyzer, 3 is the sentence analyzer, 4 is the structure processor, 5 is the display for the user and 6 is the separate analysis results search part.

/511

Figure 1 Fundamental Structure of this Invention

```
1  document input
2  sentence structure analyzer
   sentence          document structure information
3  sentence analyzer
4  structure processor
6  separate analysis results search part on the sentence analyzer
5  display for the user
```

Figure 2 Structure Processor

```
Document Input
41 structural word string analysis
   word string
42 structure processor
constructed sentence
```

Figure 3

```
Summary of the Processing
  Transitive verb form + particle "te"
  Conclusive verb form + "." + "and"
```

Example 1

```
Erase the file to finish processing.
File      noun
"wo"      particle
erase     transitive verb      irregular conjunction
"te"      particle
processing noun
"wo"      particle
finish    transitive verb
.         mark
Erase the file. And processing is finished.
```

Example 2

```
Delete the file to finish processing.
File      noun
"wo"      particle
delete     transitive verb      irregular conjunction
"te"      particle
processing noun
"wo"      particle
finish     transitive verb
.         mark
Delete the file. And processing is finished.
```


Figure 4 Structure Processor

Input sentence
43 form analysis
 word string
44 scanning specific words and assigning evaluation value
 evaluation value
46 structure processor controller
 word string
45 structure processor
 on/off control signal
output

Figure 5 Processing Example

Processing Summary
 There are more than 3 predicates.
 Analyze the sentence.
(1) Erase the file to finish processing. (2 predicates)
 Erase the file to finish processing. (unchanged)
(2) Extract the data and delete the file to finish processing. (3
 predicates)
Data noun
"wo" particle
extract transitive verb irregular conjunction
 mark
file noun
"wo" particle
delete transitive verb irregular conjunction
"te" particle
processing noun
"wo" particle
finished transitive verb past tense
 mark
Extract the data. Delete the file. And processing is finished.

Figure 6 Structure Processor

Document Input
43 form analysis
 word string
47 sentence structure analyzer
 sentence structure
48 structure processing unit
 Constructed sentence

Figure 8 Example of character string output as the final result

Processing Example
The user deletes the file to finish processing.
The user deletes the file. And he finishes processing.

Figure 9 Example of word string output as the final result

Processing Example
The user deletes the file to finish processing.
The user deletes the file.
User noun

"ha" particle
 ' mark
 file noun
 "wo" particle
 delete transitive verb
 . mark
 And he finishes processing.
 And connector
 He noun
 "ha" particle
 , mark
 processing noun
 "wo" particle
 finishes transitive verb
 . mark

Figure 7 Example of sentence structure

(A) The user deletes the file to finish processing.

User	noun
"ha"	particle
'	mark
file	noun
"wo"	particle
delete	transitive verb
"te"	particle
processing	noun
"wo"	particle
finishes	transitive verb
.	mark

(B) The user deletes the file.

User	noun
"ha"	particle
'	mark
file	noun
"wo"	particle
delete	transitive verb
.	mark

(C) And he finishes processing.

And	connector
He	noun

"ha" particle
, mark
processing noun
"wo" particle
finishes transitive verb
. mark